PRINCIPAL COMPONENT ANALYSIS (PCA)

- PCA is a modeling technique which works in an unsupervised learning setup.
- Imagine we have a dataset where we have various features of a car as the independent feature and as the dependent variable we have the price of the car (a numerical variable) or we can have the name of the car (a categorical variable).
- However, these independent features have some underlying groups which are very similar to each other.
- Suppose if the dataset has hundreds of such features, it will become very difficult to determine these types of underlying groups as we can't observe the difference from the outside.
- To find the similarity / dissimilarity inbetween groups, certain car having very high correlation while some of these cars will have lower relation however when seen on an overall basis, these two variables shows a positive correlation.
- Positive Correlation means they are indicating similar things and Negative Correlation is opposite way.
- Normally, we have two options if we have to find relationships between variables. For example, 100 variables either make thousand of 2-D plot, which is NOT EASY. Answer is PCA, we can create a PCA plot which converts the correlations or lack of correlations among the feature into a 2D graph clustering the feature that are highly correlated to one another. In Car dataset, we might find groups of features which can then Categorize into 'Dimension of Car', 'Performance of Car', 'Power' etc.
- In other method where, feature which doesn't provide much information is dropped i.e. if we have two variables that are highly correlated, we can drop one of these variable however if the feature are not statistically independent, a single feature could therefore be representing a combination of multiple type of information by a single value.

① Subalalitha Navaneethakrishnan

PCA (Principal Component Analysis) (4 Pages)

→ It is a way of identifying patterns in a data and expressing the data in such a way to highlight the similarity & differences.

→ It is uses for dimensionality reduction.

Main AIM OF PCA –
→ Keep require dimension and remove useless dimension.

Step 1 → Get some data and plot.

x	y
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2.0	1.6
1	1.1
1.5	1.6
1.1	0.9

Mean Value

$\bar{x} = 1.81$

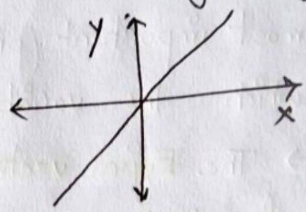
$\bar{y} = 1.91$

Step 2 → Data Adjustment →

Subtract the mean to make the data pass through origin.

$x - \bar{x}$
$y - \bar{y}$

$x-\bar{x}$	$y-\bar{y}$
0.69	0.49
−1.31	−1.21
0.39	0.99
0.09	0.29
1.29	1.09
0.49	0.79
0.19	−0.31
−0.81	−0.81
−0.31	−0.31
−0.7	−1.01



This data will have mean "0".

Step 3 → Calculate the covariance matrix.

* Covariance is a measure between 2 dimensions. They show how two variables vary together.

*
$$Cov(x,y) = \sum_{i=1}^{n} \frac{(x-\bar{x})(y-\bar{y})}{n-1}$$

Let see a small example, x : 2.1, 2.5, 3.6, 4.0
y : 8, 10, 12, 14.

$\bar{x} = 3.1$
$\bar{y} = 11$

$$Cov(x,y) = \frac{(2.1-3.1)(8-11) + (2.5-3.1)(10-11) + (3.6-3.1)(12-11) + (4.0-3.1)(14-11)}{3} \quad (n=4,\ 4-1=3)$$

$= 2.26$ → +ve value, x and y value varies positively.

- So, if covariance is +ve, x and y varies positively. If x increases then y also increases (or vice versa)
- So, if covariance is −ve, x and y varies negatively. So, if x increases then y decreases.

So what is covariance matrix?

	x	y
x	cov(x,x)	cov(x,y)
y	cov(y,x)	cov(y,y)

So, if there are two variable x and y.
we will have 2x2 covariance matrix.

Covariance matrix for Data Considered is given as

$$\text{cov} = \begin{pmatrix} 0.616 & 0.6154 \\ 0.615 & 0.716 \end{pmatrix}$$

→ Since the non-diagonal elements in this covariance are +ve, we can expect that both $x+y$ varies increases together.

Step 5 → Calculate the eigen vector and eigen values for the covariance matrix.

$$\text{eigen values} = \begin{pmatrix} 0.4908 \\ 1.25402 \end{pmatrix} \quad \text{eigen vectors} = \begin{pmatrix} -0.735 & -0.678 \\ 0.677 & -0.731 \end{pmatrix}$$

The most important (principal) eigen vector would have the direction in which the variable are strongly correlated.

Step 6 → The Eigen vector with highest Eigen value will be choosen for PCA.

* Now we can ignore the other dimension.

n-dimensions of Data (features) → n Eigen Vector → where P < N. Choose P eigen Vectors.

In our case x and y — 2-Eigen vector — Hence dimensionality reduction.

So now,

The final Data = Row Feature Vector X Row data adjust.

Row feature vector → It is the matrix with the eigen vectors in the column transposed, so that they are now in rows.

Row Data Adjust → It is the mean-adjusted data transpose (i.e,) the data items are in each ~~rows~~ columns. with each row holding a seperate dimension.

* The final data is the final dataset, with data items in columns and dimension's along rows.
* our data had the two axis x and y. So our data was in terms of Row data adjust & vectors.
* Now they are in terms of eigen vector (principal component).

③ - The new data set would have reduce dimensionality, its dimension, if we have choosen to cut an eigen vector.

- The other transformation we can make is by taking only the eigen vector with highest eigen value (Dimensionality reduction Again).

Bringing back Data back →

Row Original Data = (Row Feature Vector^T × Final Data) + Original Mean.

—x—

Eigen Vector of a Matrix

$A_x = \lambda x$ — (1)

A → Given matrix

λ → Eigen value

x → eigen vector.

x represent eigen vector (non-zero).

λ is a eigen value.

Let's find eigen value and eigen vectors

Step 1 → $A = \begin{pmatrix} 5 & -3 \\ -6 & 2 \end{pmatrix}$.

Each column represent a vector.

So, in short we strech a matrix by λ.

Step 2 → $Ax = \lambda x_2$ — (1)

$Ax - I\lambda x = 0$. —(2) where $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ → Identity matrix

Step 3 → $(A - I\lambda)x = 0$ —(3)

determinate $|A - I\lambda| = 0$ and $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Step 4 → $\begin{vmatrix} 5-\lambda & -3 \\ -6 & 2-\lambda \end{vmatrix} = 0$.

Step 5 → $(5-\lambda)(2-\lambda) - 18 = 0$.

Step 6 → $\lambda^2 - 7\lambda + 10 - 18 = 0$.

Step 7 → $\lambda^2 - 7\lambda - 8 = 0$.

Step 8 → $(\lambda + 1)(\lambda - 8) = 0$.

eigen values are $\lambda = -1$ and $\lambda = 8$. ← Step - 9.

So, eigen valus are ← Step 10

$\lambda = -1$ and $\lambda = 8$.

From eqn (3).

$(A - I\lambda)(x) = 0$. ← Step 11

Considr $\lambda = -1$. ← Step 12

$\begin{pmatrix} 5-\lambda & -3 \\ -6 & 2-\lambda \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$\begin{pmatrix} 6 & -3 \\ -6 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

$6x_1 - 3x_2 = 0$. ← Step 13

$-6x_1 + 3x_2 = 0$.

$2x_1 - x_2 = 0 \Rightarrow x_2 = 2x_1$.

Therefore, eigen value of $\lambda_1 = -1$ is $\frac{1}{2}$. ← Step 14

cross check of above ans, from eqn (1), $Ax = \lambda x$.

$Ax = \begin{pmatrix} 5 & -3 \\ -6 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ -2 \end{pmatrix}$ i.e, $(5\times1)-(-3\times2)$ / $(-6\times1)-(-2\times2)$ $= -1\begin{pmatrix} 1 \\ 2 \end{pmatrix}$

Matrix multiplication

$\uparrow \lambda$ $\uparrow x$.

So, considr $\lambda = 8$.

$\begin{pmatrix} 5-\lambda & -3 \\ -6 & 2-\lambda \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $\lambda = 8$.

$-3x_1 - 3x_2 = 0$

$-6x_1 - 6x_2 = 0$ $\Rightarrow x_2 = -x_1$

Maybe $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$

Cross check, $\begin{pmatrix} 5 & -3 \\ -6 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 8 \\ -8 \end{pmatrix} = \underset{\lambda}{\underset{\uparrow}{8}} \underset{\alpha}{\underset{\uparrow}{\begin{pmatrix} 1 \\ -1 \end{pmatrix}}}$.

In this way we find eigen vector.

—x—

→ PCA is an unsupervised algorithm.

→ Applications — Noise filtering, Visualization, Feature extraction, stock market predictions, Data analysis.

→ Goal → Identify patterns in data.
Detect the correlation in data (+/−/neutral) in order to reduce dimensionality.

→ Reduce the dimension of d-dimension dataset by projecting it onto a (K)-dimensional subspace (K < d).

→ Steps in PCA –

1. Standardize the data.
2. Obtain the eigen vectors and eigen values from the covariance matrix or correlation matrix or perform Singluar Vector Decomposition (SVD).
3. Sort eigen value in descending order and choose K eigen vectors. that corresponds to K largest eigen values where K is number of dimensions of new feature subspaces (K ≤ d).
4. Construct the projection matrix W from selected K eigen vectors.
5. Transform the original dataset X via W to obtain K dimensional feature subspace Y.

→ It learn about relationship between X and Y values.

→ Find list of principal axes.

→ Highly affected by outliers.

https://plot.ly/ipython-notebooks/principal-component-analysis/

setosa.io/ev/principal-component-analysis

Udemy.